

Přehledová práce

Statistické metody v medicíně II. – p-hodnota

Statistical methods in medicine II. – p-value

Martin Hynek^{1,2}, Jana Zvárová²

¹Gennet, s.r.o., Centrum pro fetální medicínu a reprodukční genetiku, Praha²Ústav hygieny a epidemiologie, 1. lékařská fakulta, Univerzita Karlova, Praha

Korespondenční adresa: MUDr. Martin Hynek, Gennet, s.r.o., Kostelní 9, 170 00 Praha 7, tel.: +420 222 313 000, e-mail: martin.hynek@gennet.cz

Publikováno: 7. 3. 2016 Přijato: 23. 2. 2016 Akceptováno: 25. 2. 2016
Actual Gyn 2016, 8, 24-25 ISSN 1803-9588 © 2016, Aprofema s.r.o.
Článek lze stáhnout z www.actualgyn.com



Citujte tento článek jako: Hynek M, Zvárová J. Statistické metody v medicíně II. – p-hodnota. Actual Gyn. 2016;8:24-25

P-hodnota (dosažená hladina významnosti, *p*-value, *p*-level) představuje jeden z nejčastěji používaných nástrojů při statistických analýzách v medicíně a s její pomocí je možno rozhodnout o zamítnutí nulové hypotézy na zvolené hladině významnosti. Hladina významnosti je pravděpodobnost chyby, že zamítneme platnou nulovou hypotézu. Navzdory všudypřítomnosti *p*-hodnoty v medicíně je pojem velmi často špatně chápán, používán a interpretován.

Velká část statistických analýz v medicíně zahrnuje porovnávání (způsobů léčby, metod, skupin případů, apod.), kdy číselná hodnota odpovídající tomuto porovnání se nazývá efekt. Tvrzení, že efekt je nulový, nazýváme nulovou hypotézou a je obvykle opakem toho, co chceme prokázat. Přítomnost efektu (rozdílu) představuje alternativní hypotézu, která obecně tvrdí, že efekt je nenulový. Pokud zformulujeme nulovou a alternativní hypotézu, nasbíráme data a vybereme vhodný statistický test, můžeme spočítat pravděpodobnost, s jakou bychom mohli získat pozorovaná data nebo data stejně či ještě více odporující nulové hypotéze, a to za předpokladu, že je nulová hypotéza pravdivá. Tato pravděpodobnost se nazývá *p*-hodnota nebo také dosažená hladina významnosti. Čím menší je *p*, tím neudržitelnější je nulová hypotéza.

Interpretace *p*-hodnoty není vždy jednoduchá. V případě poměrně vysokých *p*-hodnot (řekněme > 0,2) můžeme říci, že data podobná našim bychom mohli získat poměrně často, pokud je nulová hypotéza pravdivá. Proto nemůžeme vyloučit možnost, že nulová hypotéza je pravdivá. Naopak, pokud dostaneme velmi nízké *p* (řekněme < 0,001), je nulová hypotéza neudržitelná, protože data podobná našim bychom dostali jen stěží náhodou, pokud by nulová hypotéza byla platná. Nulovou hypotézu v ta-

kovém případě zamítáme. Mezi těmito dvěma extrémny leží šedá zóna nejistoty, ale obvyklým postupem je zvolit si jistou mezní hodnotu α , nazývanou hladinou významnosti, a pokud je *p*-hodnota menší než tato mez, zamítneme nulovou hypotézu ve prospěch hypotézy alternativní. Nejčastěji je za tuto mezní hodnotu α volena hodnota 0,05 (5 %). Je-li *p* menší než zvolená hladina významnosti α , označujeme rozdíl za statisticky významný. V opačném případě říkáme, že rozdíl není statisticky významný. Nicméně, než tato striktní dichotomizace na statisticky významný a statisticky nevýznamný rozdíl, je s výhodou vždy uvádět i přesnou *p*-hodnotu.

Koncept *p*-hodnoty patří k hojně kritizovaným pojmům a dokonce vedl k tomu, že někteří editoři *p*-hodnotu a testování hypotéz zcela vykárali ze svých časopisů (například Journal of Basic and Applied Social Psychology, Trafimow a Marks 2015). Příčina kritiky však plyne často ze špatného pochopení a špatného použití *p*-hodnoty. Zdůrazněme proto nejčastější omyly v chápání *p*-hodnoty.

P-hodnota není pravděpodobnost, že při zamítnutí nulové hypotézy děláme chybu ani pravděpodobnost, že nulová hypotéza je pravdivá či alternativní hypotéza je nepravdivá.

P-hodnota se počítá za předpokladu platnosti nulové hypotézy a neříká nám nic o platnosti alternativní hypotézy, která je obvykle předmětem našeho zájmu. Zamítnutí nulové hypotézy však ještě neznamená, že alternativní hypotéza musí být pravdivá.

Hladina významnosti α není určena *p*-hodnotou, ale je to čistě arbitrárně volená hodnota, kterou si má zvolit experimentátor před analýzou dat. Obvykle volená hladina

$\alpha = 0,05$ vychází z historických konvencí. Je dlouhodobě argumentováno, že takto arbitrárně stanovená mez zpomaluje vědecký pokrok a má škodlivý vliv na odbornou literaturu (Johnstone 1986, Cohen 1994, Anderson 2000, Berkson 2006, Smedslund 2006, Hubbard a Lindsey 2008). Příkladem může být publikační bias ve prospěch studií se signifikantním závěrem (Altman 1991), což ale není chyba zvolené hladiny významnosti, ale vydavatelů vědeckých časopisů, kteří statisticky nevýznamné výsledky nepublikují.

Konečně p -hodnota nám neříká nic o velikosti pozorovaného efektu, nelze ji tedy použít pro porovnávání, ale je čistě vodítkem udržitelnosti či neudržitelnosti nulové hypotézy.

Závěrem uvádíme čtyři klíčové body pro správné použití p -hodnoty: testování hypotéz a p -hodnota jsou užitečné nástroje a poskytují důležité informace, ale vždy *dohromady s dalšími nástroji* (například deskriptivní statistika, správně navržený experiment či studie, ověření vhodnosti použití zvolené statistické metody na nasbíraná data), je nezbytná *správná interpretace výsledků*, statisticky významné výsledky musí být vždy potvrzeny z *opakovaných studií*, abychom jim mohli věřit a konečně abychom *zhodnotili kvalitu studie*, je nutné se dívat na všechny její dílčí kroky a ne jen na vlastní p -hodnotu. Kdykoliv je to možné, doporučujeme použít raději interval spolehlivosti při interpretaci odhadu velikosti zkoumaného efektu než p -hodnotu.